When are our forecasts ahead better, i.e., more accurate?

Konner Macias
004603916

1 Introduction

The main question we are trying to address in this paper is: **When are our forecasts ahead better, i.e., more accurate?**

(a) When the only information we use is the historical information of the time series we want to forecast, and nothing else, as when using SARIMA(p,d,q)(P,D,Q) or perhaps SARIMA with garch?

(b) When the variable we want to forecast is the dependent variable in a traditional regression model and we use other information in the form of other variables that play the role of exogenous variables? Such model could also have dummies for seasonals and polynomial trends for trend. The residuals of such a model probably need to be modeled as AR or ARIMA model to account for autocorrelation. This is the approach we used when we studied gls to approach regression with autocorrelated residuals.

(c) When we use information in other variables, but the variable we want to forecast is a dependent and independent variable at the same time, and so are the other variables both dependent and independent, as in VAR?

(d) When we average the forecasted values of (a), (b), (c), to obtain a consensus forecast.

These questions always arise in practice in a number of sciences, among them economics and meteorology. Nobody would ever use a single model to make a forecast for time $t+1$. Almost all areas that predict the future fit several models, and then they average the forecasts obtained for each future time $t+1$ from all the models to obtain what they called the consensus forecast for time t. Google the word consensus forecast to learn more. The Blue Chip consensus forecasts of economic indicators are an example of how the many forecasting companies out there combine their forecasts into a consensus forecast.

2 Data

In this section, we will describe that data that we will use to answer the question posed in the introduction. The variable that we will want to forecast is housing starts in the United States. Housing starts is considered to be a leading indicator of what might come next in the economy. If housing construction starts to flourish that should be an indication of prosperity to come, some economists think. Other information that will be used for the models that use the other variables is unemployment rate, and unemployment rate for women. The source of the data is FRED (Federal Reserve Economic Data, `https://fred.stlouisfed.org`). The variables are observed from January 1st 1959 to August 1st 2018. We will use January 1st 1959 to August 1st 2017 as training data, to fit the model, and then we will forecast from September 1st 2017 to August 1st 2018. Below is a table which summarizes this information about the training and testing sets.

Variable Name	R Name	Description and Source	Training Set	Test Set
HOUSTNSA	hs	Housing Starts: Total new privately owned housing units (in thousands). Monthly. Not seasonally adjusted **This is the variable we want to forecast**	Jan 1, 1959 Aug 1, 2017	Sept 1, 2017 Aug 1, 2018
LNU04000002	uw	Unemployment rate: Women (percent) Monthly. Not seasonally adjusted	Jan 1, 1959 Aug 1, 2017	
UNRATENSA	ur	Civilian Unemployment Rate (percent) Monthly. Not seasonally adjusted	Jan 1, 1959 Aug 1, 2017	

3 Data description, unit roots test, volatility checking and cointegration

3.1 Decomposition and Seasonal Box Plot Analysis

Here we examine all components of the time series (observed values, trend component, seasonal component, and random component)

Figure 1. Decomposition of `hs`
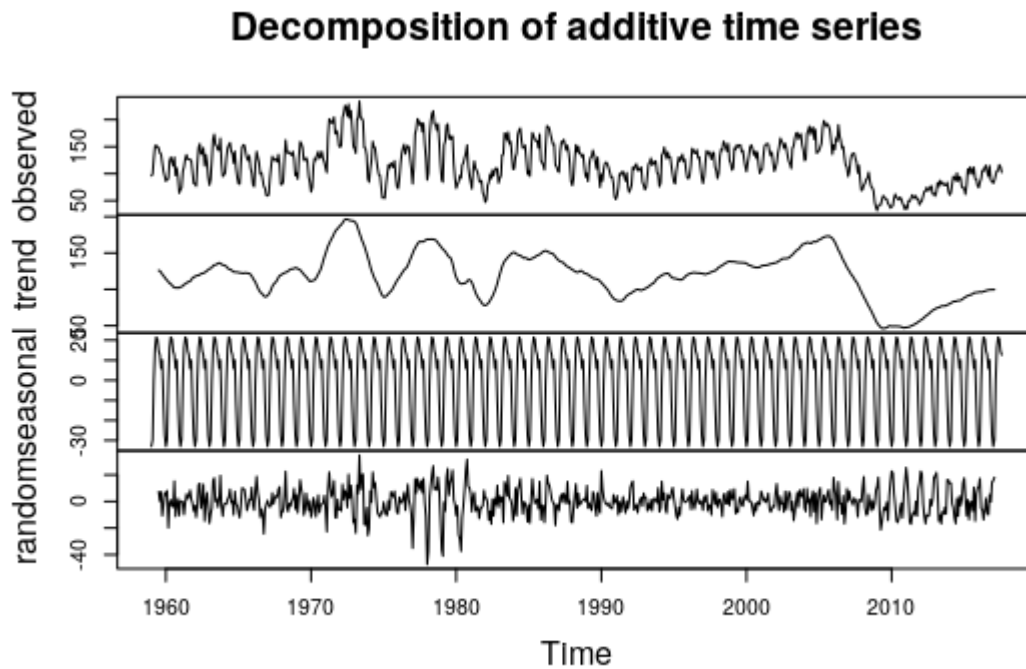
Looking at Figure 1, first within the observed plot, we find it difficult to depict an overall trend. This is justified by the trend component of the plot. Interestingly, the trend appears stationary until mid 2000's where it took a massive drop, most likely due to the 2008 housing crisis[1]. From the seasonal component, we notice a perfectly repeating pattern. The random component overall appears to be stationary with no obvious trends, similar to white noise. Next, let's examine the decomposition of unemployment rate for women (`uw`).

Figure 2. Decomposition of `uw`

[1]https://www.americanprogress.org/issues/economy/reports/2017/04/13/430424/2008-housing-crisis/
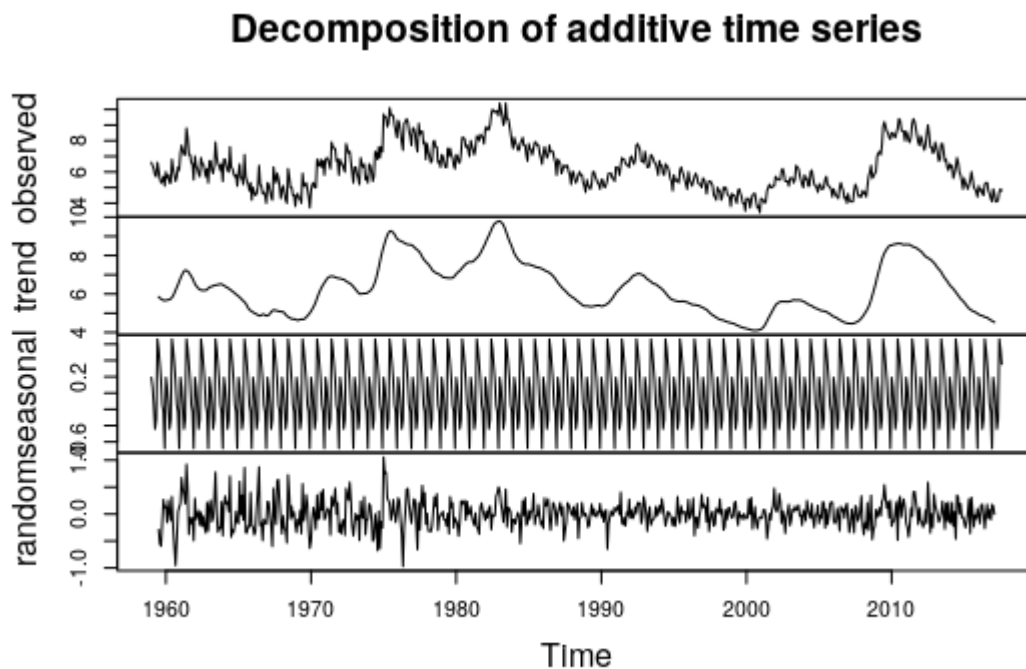
Looking at Figure 2, first within the observed plot, we notice periods of slow decay of unemployment followed by a sudden and quick rise. We notice that the greatest rise in terms of magnitude occured around 2008. The time plot of the data does appear to be stationary. The trend component is not regular, and only slightly indicates some stationarity. The seasonal component is obvious with a repeating patttern occuring each year. The random component appears stationary. Next, let's examine the decomposition of the overall Civilian unemployment rate (ur).

Figure 3. Decomposition of ur
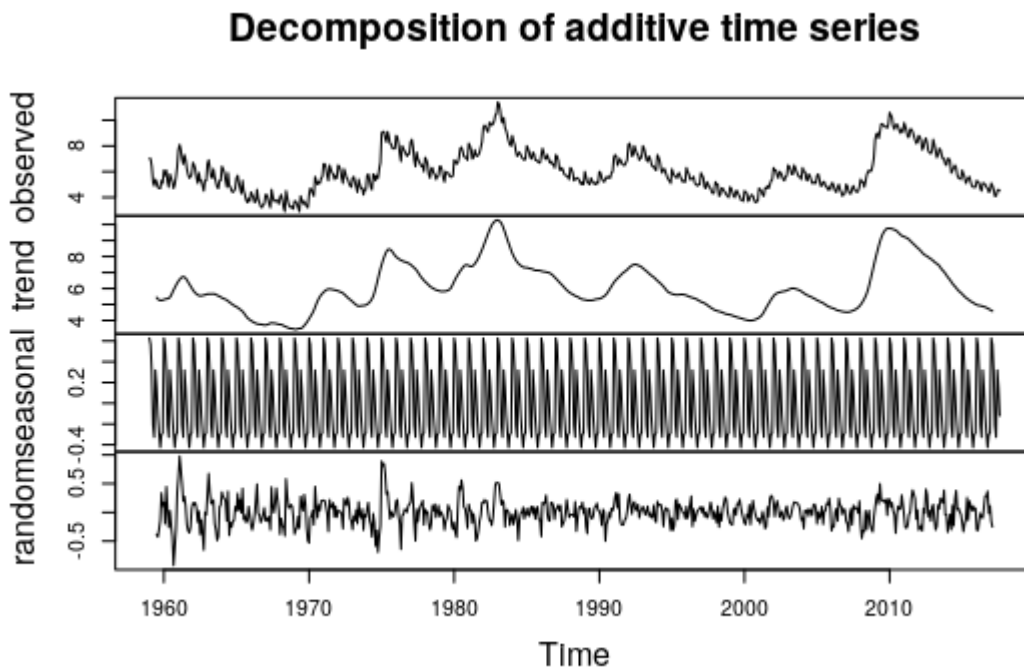
At first glance of Figure 3, we notice a near identical pattern as the unemployment rate for women, as it would intuitively suggest. All comments made for Figure 2 in terms of trend, seasonality, and stationarity of the random component holds. The only observable difference appears in the overall magnitude of the fluctuation in the observed component, as this does not fluctuate as great as the observed component for ur.
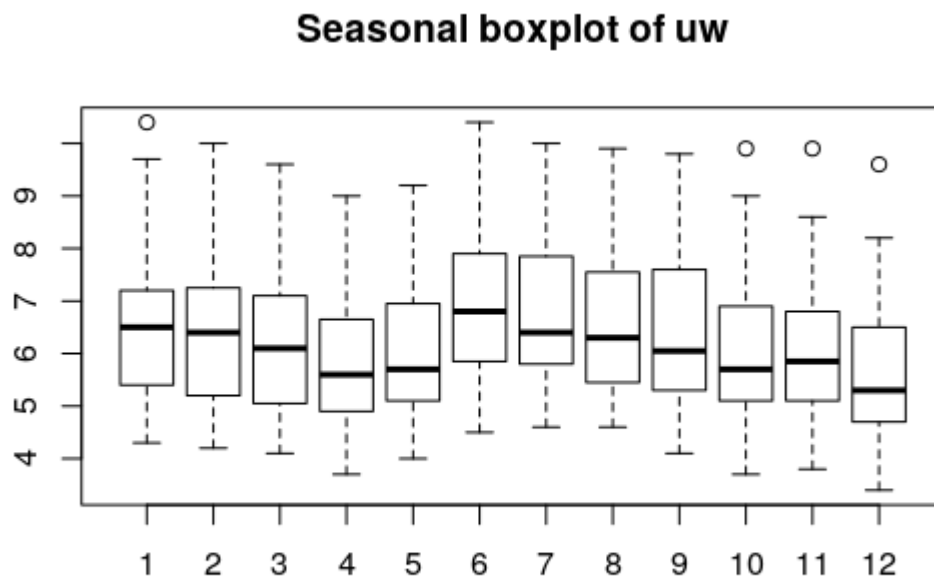
Let's now examine the seasonal boxplots of each variable we are inspecting.

Figure 4.

Seasonal boxplot of hs
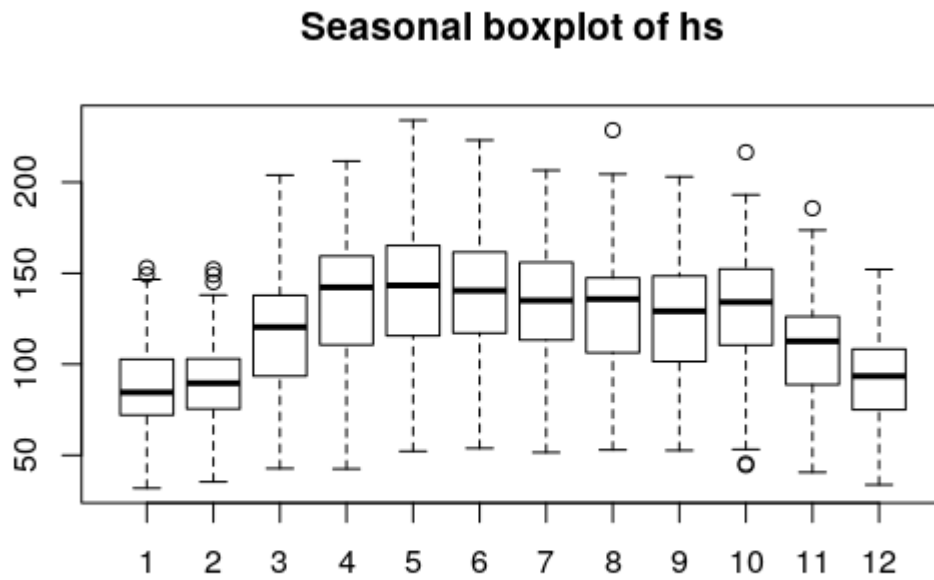
We notice in Figure 4 that the number of housing starts is greatest outside of the winter months. However, we notice that the IQR of these spring and summer months is very large indicating that it can very well fluctuate.
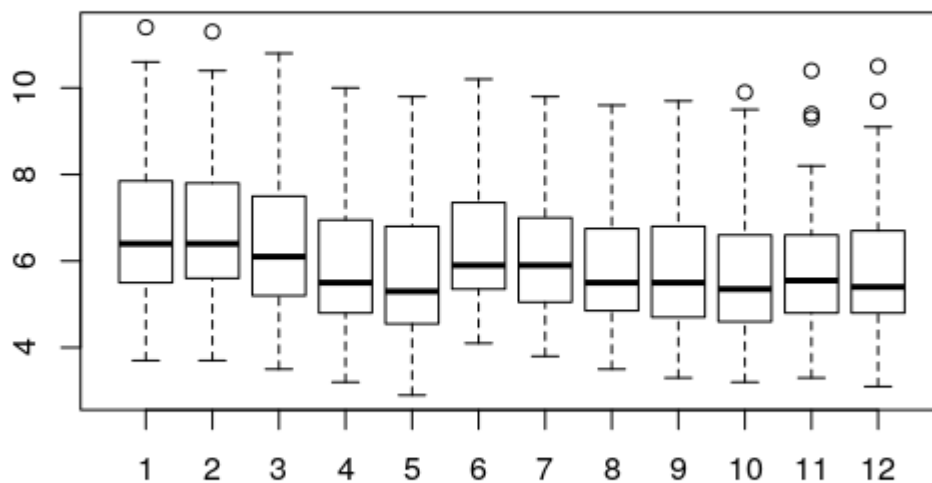
Figure 5.

Seasonal boxplot of uw

We notice in Figure 5 that the unemployment rate for women stays relatively the same throughout the year. We two peaks at January and June, with gradual decays occuring after those months.

Figure 6.

Seasonal boxplot of ur

We notice a near identical pattern for civilian umployment rate as with women's umeployment rate. There is a slight decay after January and June.

3.2   Unit Root, Cointegration, and Volatility Tests

Before we begin to model, we must make sure we are not violating any underlying assumptions.

Unit Root Tests
To help to guarantee we are not getting a spurious regression, we perform the augumented Dickey-Fuller test which has a null hypothesis stating that a given time series x_t is a random walk.

We obtain p-values of $0.2943, 0.119, 0.09635$ for housing starts, unemployment rate for women, and civilian umeployment rate, respectively. Thus, we fail to reject the null hypothesis for all, meaning that these three time series are random walks.

Cointegration Tests
To verify that two series are not cointegrated and do not share an underlying stochastic trend, we perform the Phillips-Ouliaris test. The null hypothesis for this test is that the two series it is considering are NOT cointegrated.

When comparing between each pair possible across the variables we are considering, we obtain p-values less than 0.01, indicating that each of the variables are cointegrated with each other and share a common stochastic trend. We must take this into account when modeling, especially for the upcoming VAR model.

Volatility Tests
A heteroskedastic or volatile series exhibit random periods of increased variance, meaning that the variance is correlated in time even though a series may look like white noise. If this is the case for our variables, we must model the variance appropriately. Upon examination of the time plots of each of the series, we do not notice these outburst in volatility. Here, I examine the acf the data subtracted by its mean and the acf of the data subtracted by its mean squared with Figures 7 and 8.

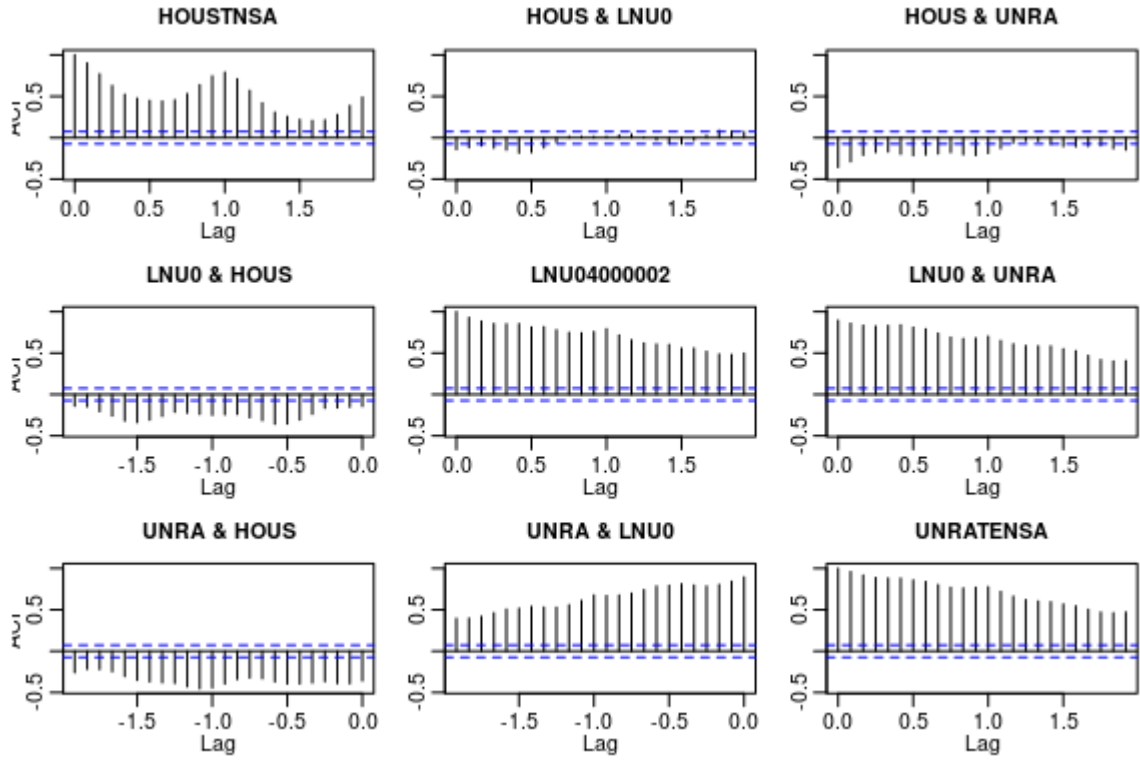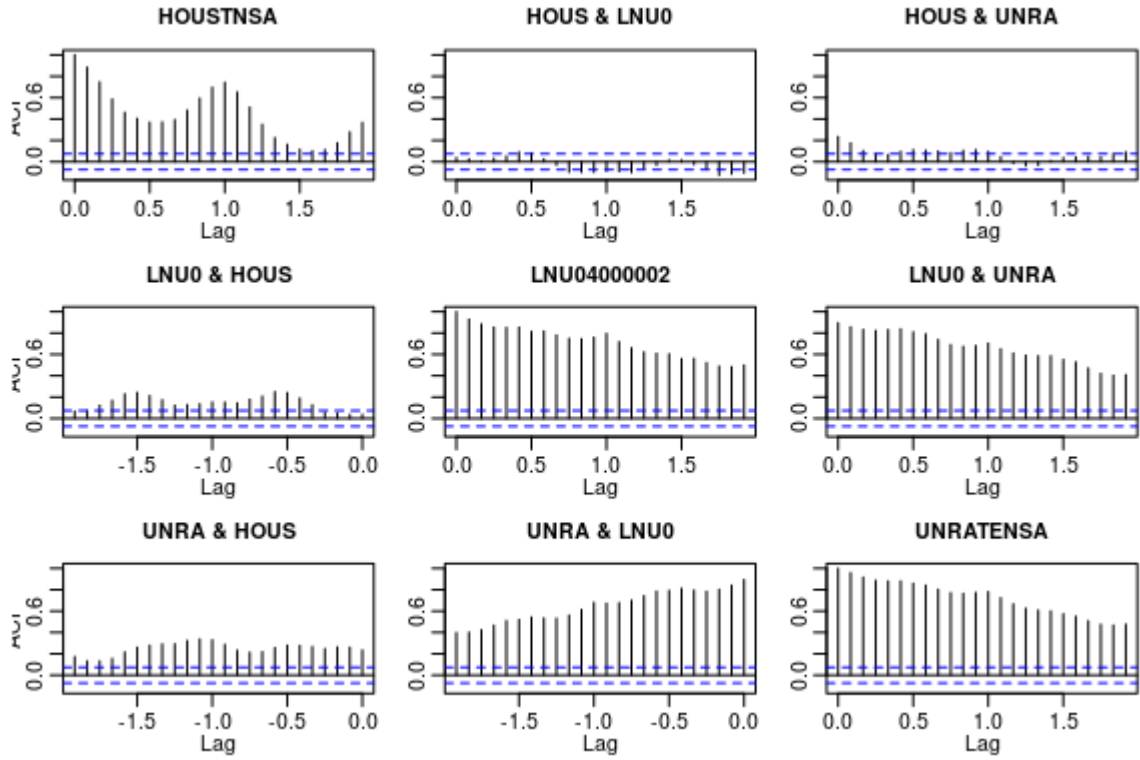**Figure 7. ACF of (data - mean(data))**

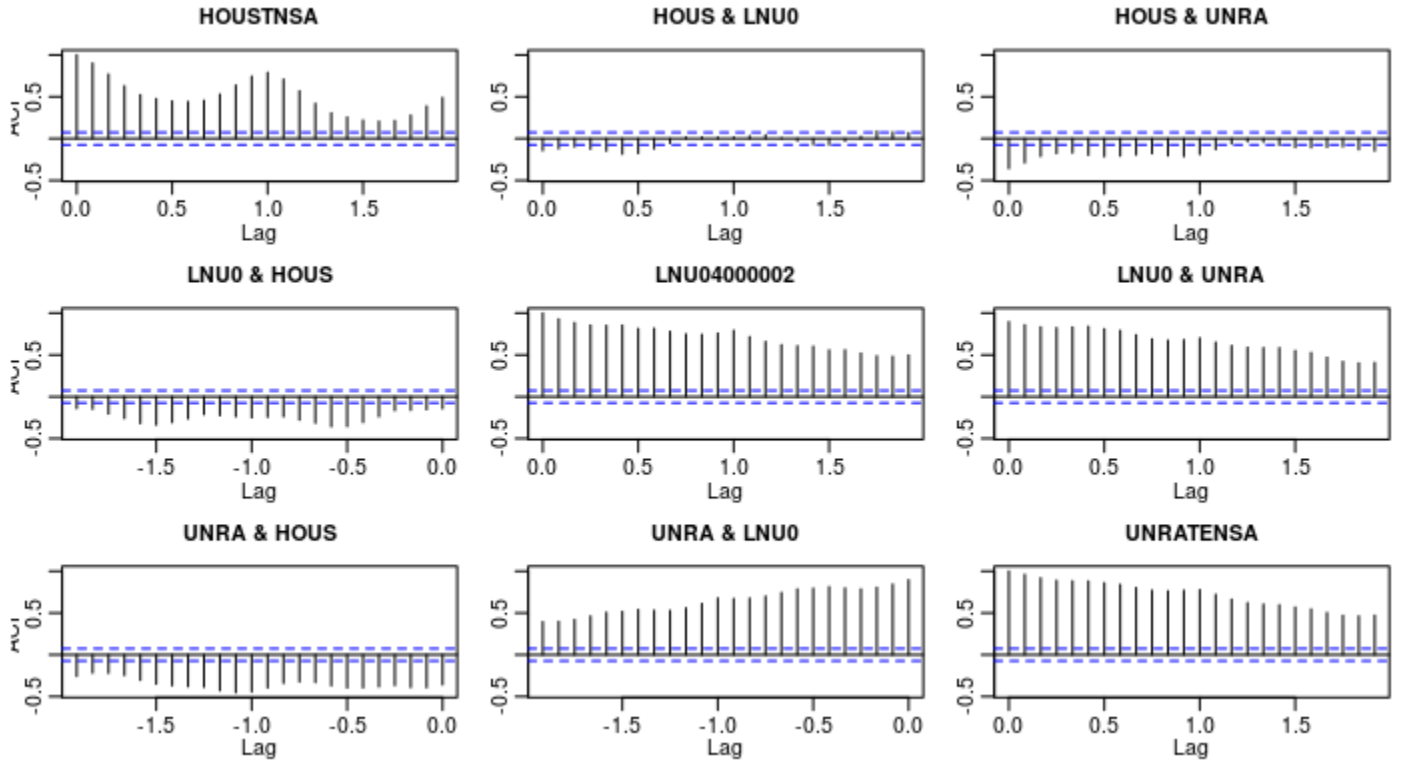**Figure 8. ACF of (data - mean(data))^2**

Looking at Figures 7 and 8, we notice no volatility as the spikes appears to come in a periodic fashion with no drastic outbursts.

3.3 Autocorrelations and Cross-Correlations

Here we wish to better understand the relationships between the variables.

Figure 9. ACFs and CCFs of the Data

As indicated by Figure 7, we notice several significant lags when comparing in the cross-correlograms of housing starts vs. unemployment rate for women and housing starts vs. civilian unemployment rate. This indicates that unemployment rate of women and civilian unemployment affect housing starts. In fact, when comparing across each cross-correlogram, we notice that each variable affects each other as all have significant lags. Looking at the identity diagonal, we notice that we will later have to difference the data.

## 4 SARIMA(p,d,q)(P,D,Q) perhaps with garch model?

In this section, I fit several SARIMA models to the training set of housing starts only. Before I begin, I must follow the standard procedure for appropriately identifying and fitting an ARIMA model. In section 3, we argued that the data was not volatile, thus there would be no need to run a potential GARCH with SARIMA model. Instead, a traditional SARIMA should handle the problem appropriately.

Transformation
Here we check whether we can transform the variance to make it more stationary allowing for us to better model the data.

Figure 10.

Org. vs. Log vs. Sqrt Transform
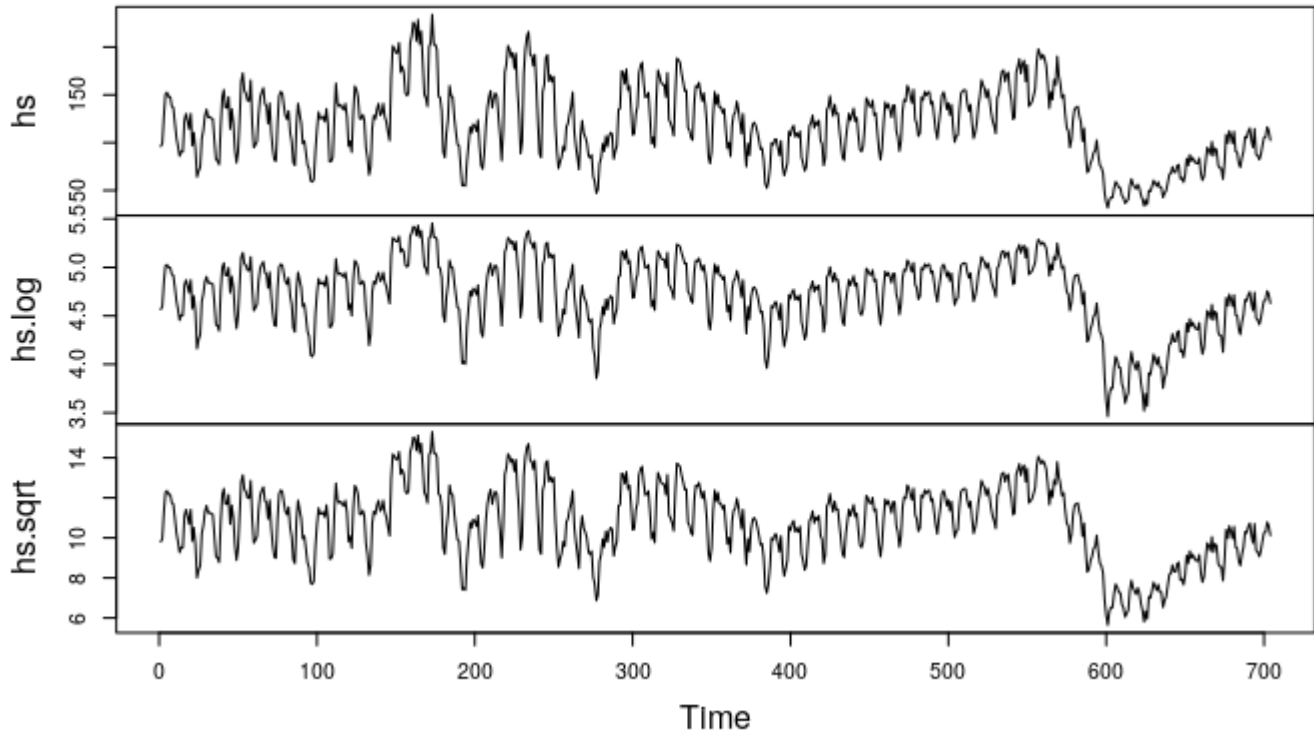
Examining Figure 10, we focus on the consistency of the magnitude of fluctuation across the entire time plot. The most consistent appears to be the `log` transformation, though the `sqrt` is close as well. We choose the log transformation over the sqrt due to its more general nature. Overall, looking at the time plot, we notice stationarity until the mid 2000s when housing starts hit a drastic decline. We call the new transformed data $hs^\star$.

Differencing

Now we must examine whether differencing must be done to the data to make it more stationary.

Figure 11.
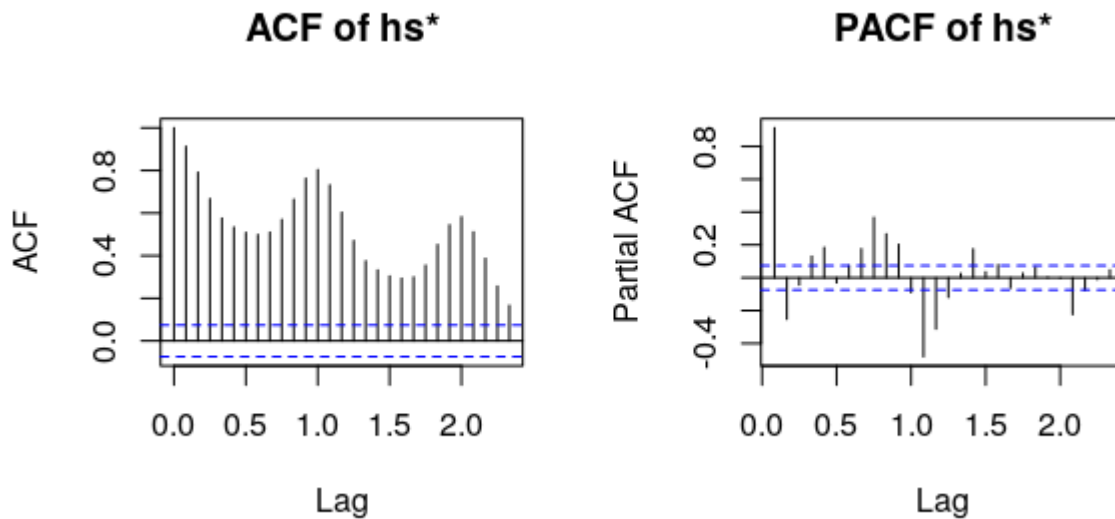
Figure 11 contains all statistically significant lags, but since the data itself is non-stationary, we cannot extract

8

much information. The lags there appear to rise and falll in a pattern potentially indicating some seasonality. Next, we attempt to regularly difference.
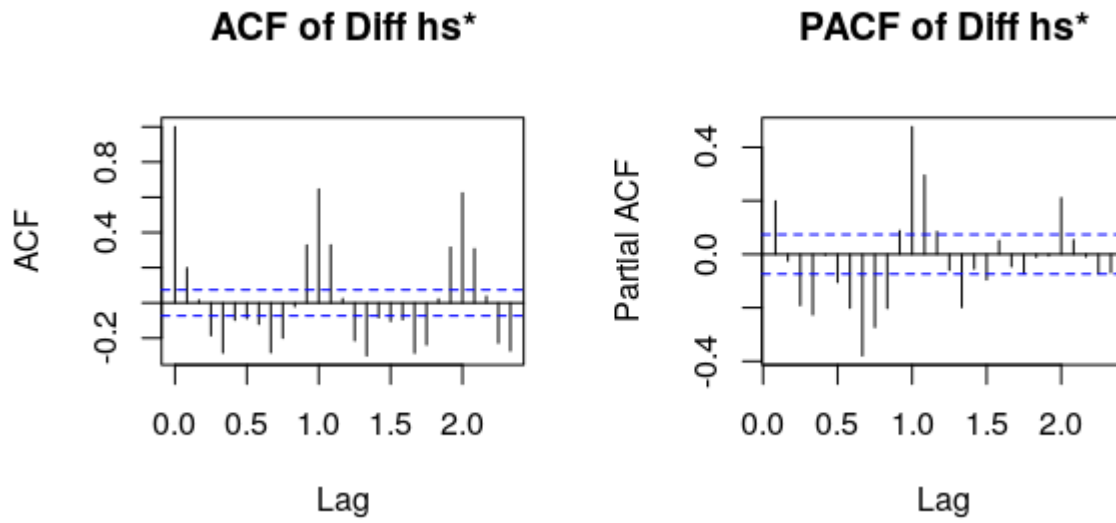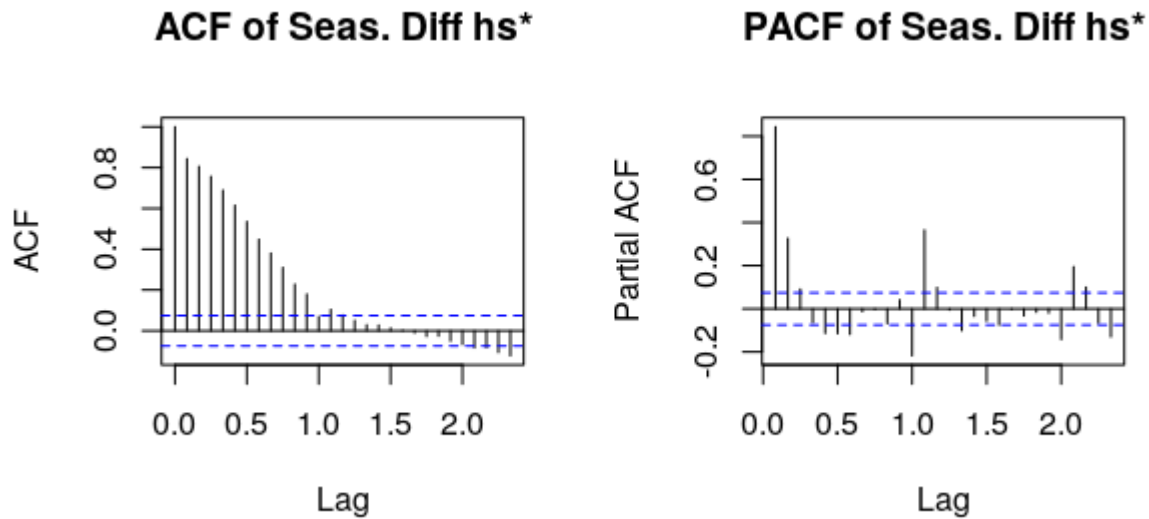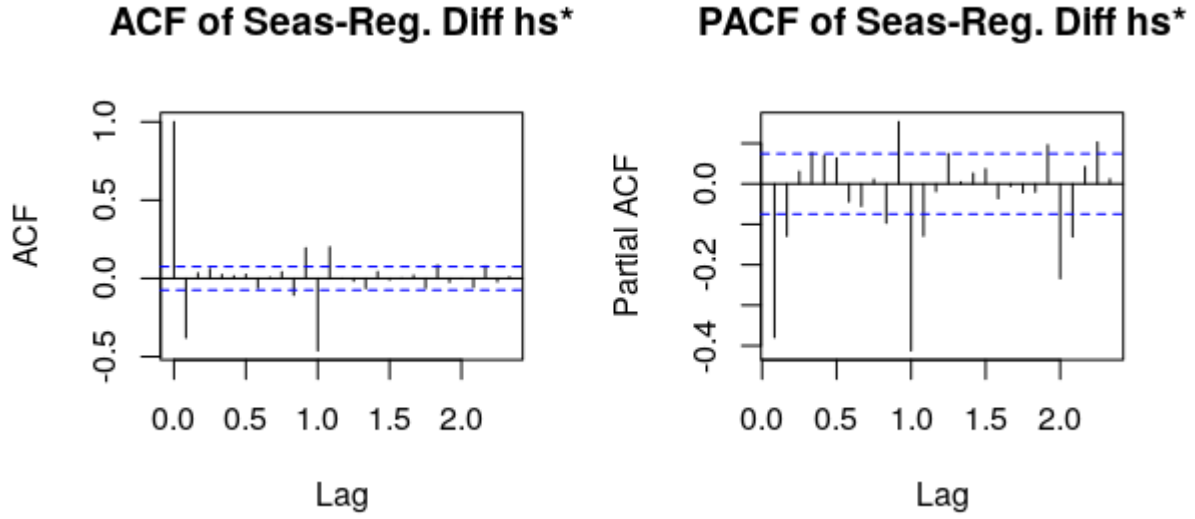
Figure 12.

ACF of Diff hs*

PACF of Diff hs*

Figure 12 illustrates the results of regularly differencing the data. We notice that the data is still not entirely stationary as there are still several significant lags, each occuring around the end of year also indicating some underlying seasonality. Next, we take a look at solely seasonally differenced data.

Figure 13.

ACF of Seas. Diff hs*

PACF of Seas. Diff hs*

Looking at Figure 13, we notice that data still is not stationary yet. There appears to be a slow decay of lags in the ACF from the positive to negative direction, possibly indicating an underlying trend in the data.

Figure 14.

ACF of Seas-Reg. Diff hs*   **PACF of Seas-Reg. Diff hs***

Lastly, looking at Figure 14, we see that compared to the other ACFs, seasonally differencing the regularly differenced data proved to be the most suitable method for making the data stationary. There are still a few statistically significant lags, but this is acceptable for now as they tend to occur at the seasonal time steps. We can call this new data $hs^{\star\star}$. In combination of examination of the PACF and ACF, we obtain hints of a potential AR and MA lurking within the data. Next, we explore this further.

Identifying a SARIMA Model
For the sake of totality, I consider many possible combinations of models for a given SARIMA model, though we can intuitively eliminate some of the following models.
We consider the following SARIMA models to best fit the data.

$$\text{Model 1: } SARIMA(0,1,0)(0,1,0)_12$$
$$\text{Model 2: } SARIMA(0,1,0)(1,1,0)_12$$
$$\text{Model 3: } SARIMA(0,1,0)(0,1,1)_12$$
$$\text{Model 4: } SARIMA(0,1,0)(1,1,1)_12$$
$$\text{Model 5: } SARIMA(0,1,1)(0,1,0)_12$$
$$\text{Model 6: } SARIMA(0,1,1)(1,1,0)_12$$
$$\text{Model 7: } SARIMA(0,1,1)(0,1,1)_12$$
$$\text{Model 8: } SARIMA(0,1,1)(1,1,1)_12$$
$$\text{Model 9: } SARIMA(1,1,0)(0,1,0)_12$$
$$\text{Model 10: } SARIMA(1,1,0)(1,1,0)_12$$
$$\text{Model 11: } SARIMA(1,1,0)(0,1,1)_12$$
$$\text{Model 12: } SARIMA(1,1,0)(1,1,1)_12$$
$$\text{Model 13: } SARIMA(1,1,1)(0,1,0)_12$$
$$\text{Model 14: } SARIMA(1,1,1)(1,1,0)_12$$
$$\text{Model 15: } SARIMA(1,1,1)(0,1,1)_12$$
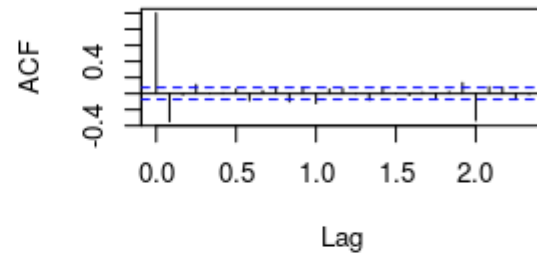$$\text{Model 16: } SARIMA(1,1,1)(1,1,1)_12$$

Let's examine the plots of their residuals.

Figure 15.

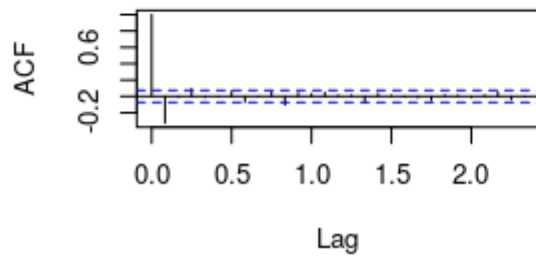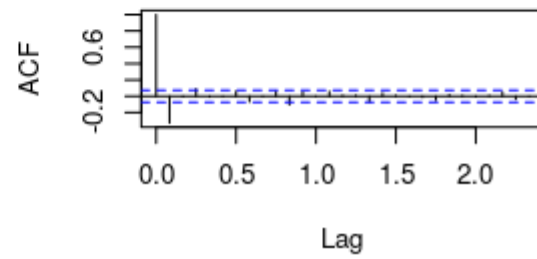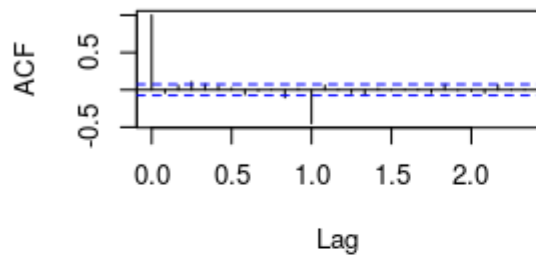## ACF of Res of M9
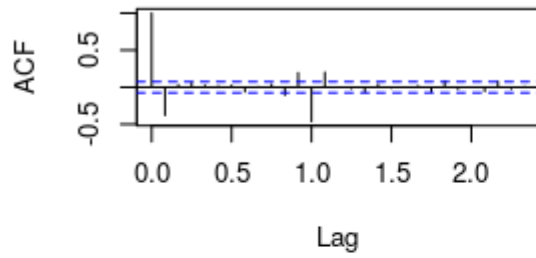
ACF of Res of M10

ACF of Res of M11

ACF of Res of M12

ACF of Res of M13
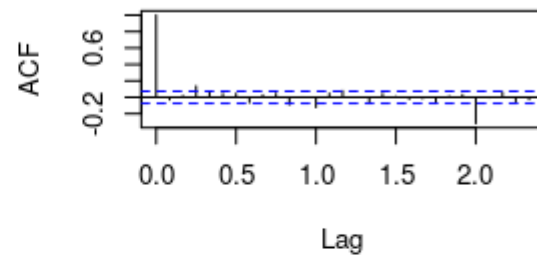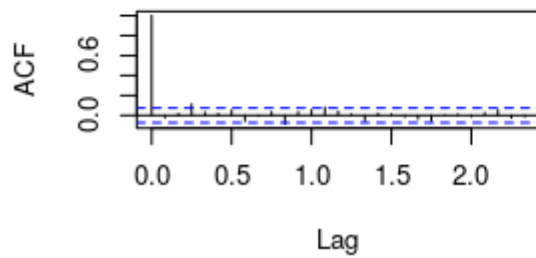

ACF of Res of M14

ACF of Res of M15

ACF of Res of M16
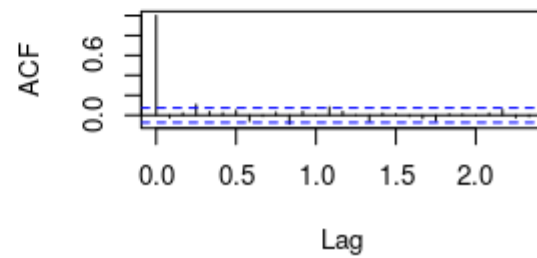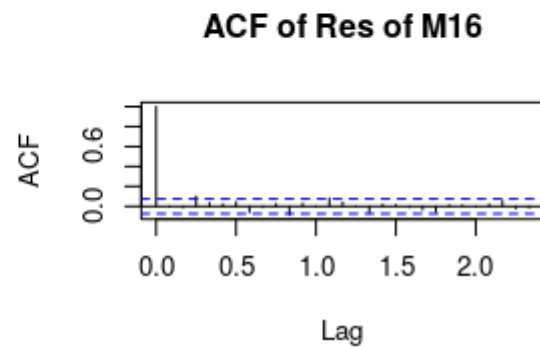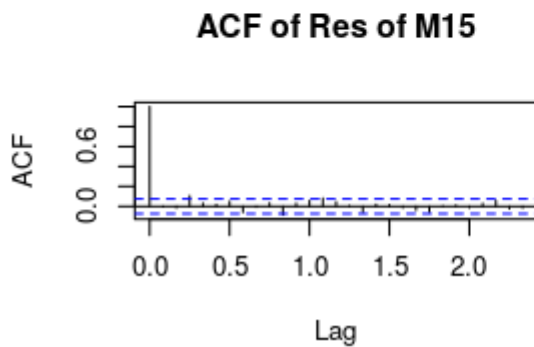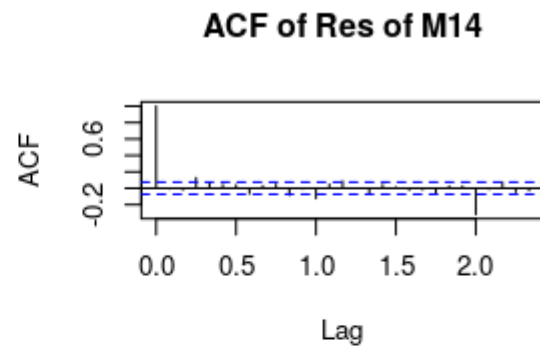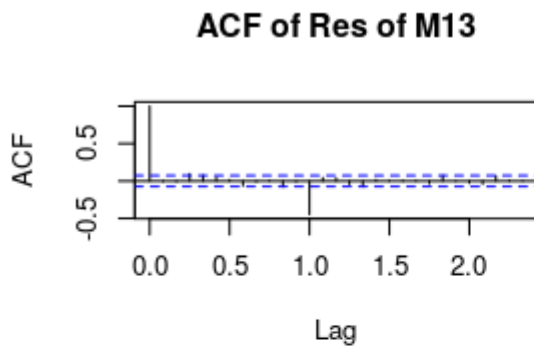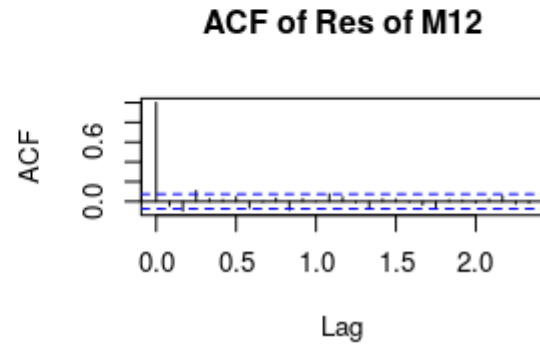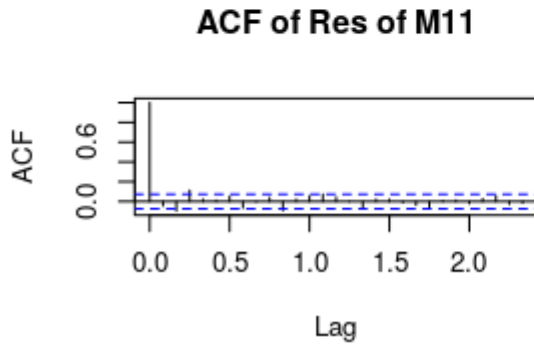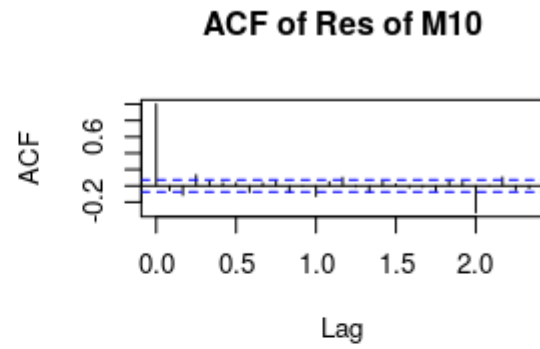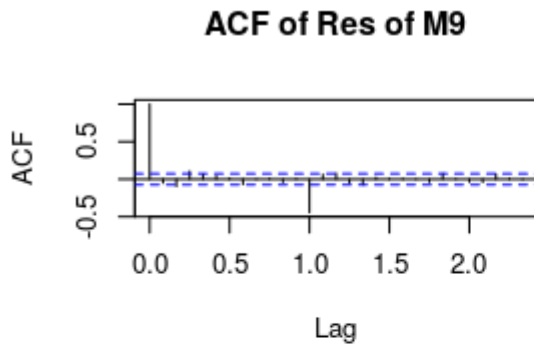
Here I decide to remove models 1,2,3,4,5,6,9,10,13, and 14 as they all contain either early statistically significant lags or additional lags around the seasonal time steps indicating that they do not do an adequate job of correctly modeling the data.

Next, I compare the p-values of the coefficients of the remaining subset of models to check whether they are significant or not. I select against those whose model coefficients are all non-significant. With this condition of ALL being non-significant, I am unable to erase any of the remaining models from consideration. Here are my

remaining models.

$$\text{Model 7: } SARIMA(0,1,1)(0,1,1)_12$$
$$\text{Model 8: } SARIMA(0,1,1)(1,1,1)_12$$
$$\text{Model 11: } SARIMA(1,1,0)(0,1,1)_12$$
$$\text{Model 12: } SARIMA(1,1,0)(1,1,1)_12$$
$$\text{Model 15: } SARIMA(1,1,1)(0,1,1)_12$$
$$\text{Model 16: } SARIMA(1,1,1)(1,1,1)_12$$

Let's now compare against the remaining models against several different criteria.

Figure 16.

| Model	AIC	$\hat{\sigma^2}$	RMSE
$SARIMA(0,1,1)(0,1,1)_12$	-1342.67	0.008094	9.223459
$SARIMA(0,1,1)(1,1,1)_12$	-1343.46	0.008061	9.371128
$SARIMA(1,1,0)(0,1,1)_12$	-1342.67	0.008094	9.223459
$SARIMA(1,1,0)(1,1,1)_12$	-1343.46	0.008061	9.371128
$SARIMA(1,1,1)(0,1,1)_12$	-1342.44	0.008076	9.195706
$SARIMA(1,1,1)(1,1,1)_12$	-1343.41	0.008042	9.35449

We see eerily similar results across the remaining models. In terms of AIC, the best model proved to be $SARIMA(0,1,1)(1,1,1)_12$ and $SARIMA(1,1,0)(1,1,1)_12$ as they had the lowest value (-1343.46) in comparison. In terms of $\hat{\sigma^2}$, the best model is $SARIMA(1,1,1)(1,1,1)_12$ since it has the lowest value (0.008042) in comparison. In terms of RMSE, the best model is $SARIMA(1,1,1)(0,1,1)_12$ as it has the lowest value (9.195706) in comparison. With results like these, one cannot make a particularly bad decicsion. I decide to select $SARIMA(1,1,1)(0,1,1)_12$ as my final model as RMSE takes precedence over the others since it is considered the most valid metric for comparing accuracy across different models. Here is the forecasting model written in polynomial form:

$$(1 + 0.1184B)(1 - B^{12})(1 - B)y_t = (1 + (-0.8843)B^{12})(1 - 0.2433B)w_t$$

We know from above that the residuals of the model are close to white noise, so we have a valid model and thus can forecast.
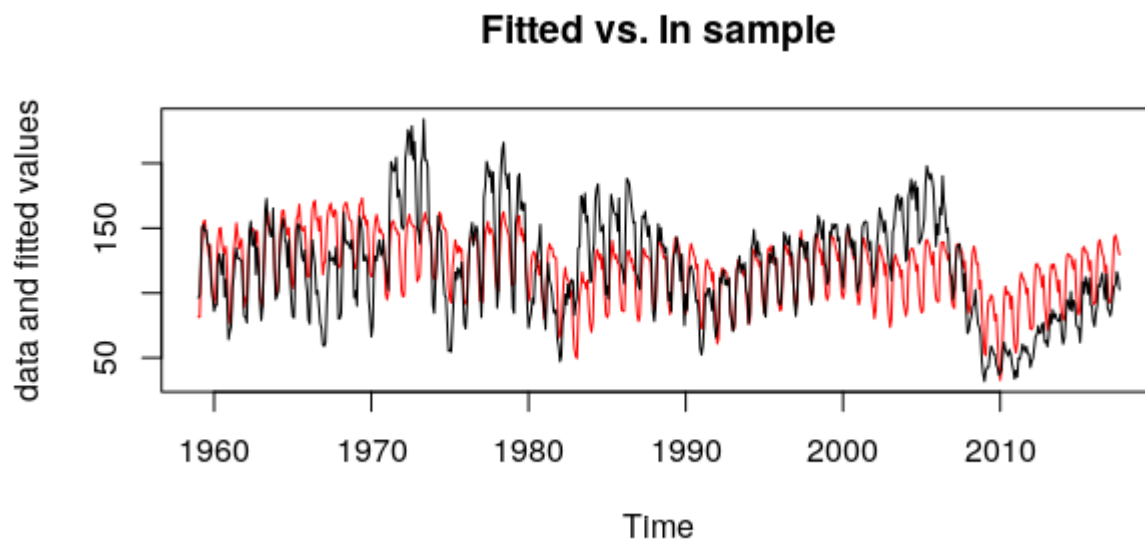
Figure 17.

Forecast of Identified Model
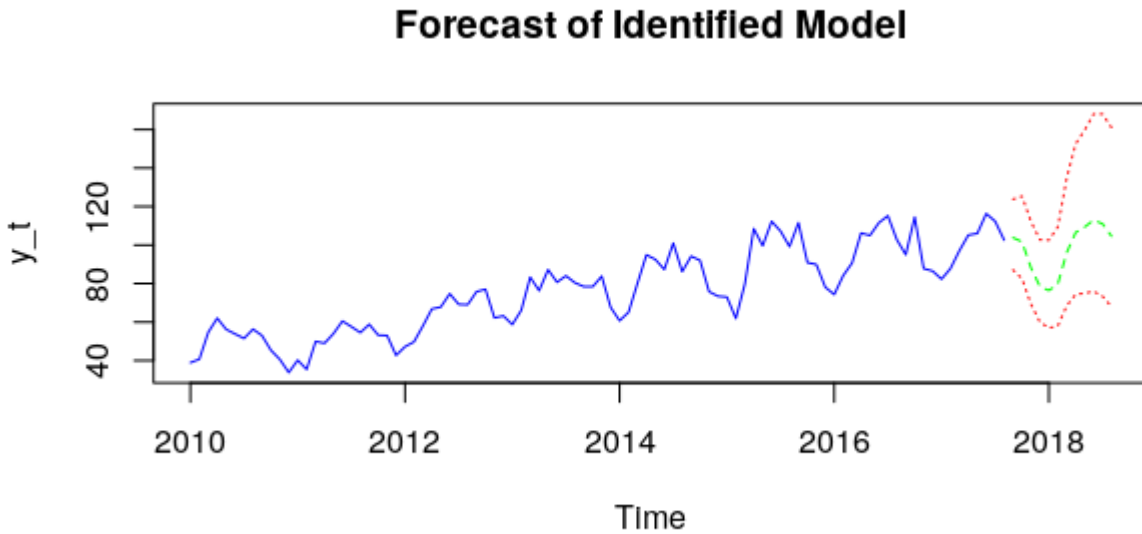
Examining Figure 12, we notice that the predictions follows a similar rise and fall pattern that we noticed in previous years' data.

5   Regression with Autocorrelated Errors

Here I perform a classical regression, or in other words, structural equation model to model the relation between housing starts and the other independent variables (`uw` and `ur`). In contrast to VAR models, classical regression models require that the dependent variable (`hs`) does not cause the independent variable. The direction of casuality goes only from the independent to the dependent variables and not the other way around.

The initial regression model of simply using only the other two variables as independent variables achieved an R^2 of 28.66%. Next, I wanted to address the seasonally depicted in the decomposition plot of housing starts in Figure 7. Here, I create a months dummy variable which can be added into our model to help add some weight on the changing values in housing starts over the year. With this new variable added in, I achieve an R^2 of 42.15%.

Figure 18.


As indicated in Figure 18, the fitted values now do a great job of rising and falling in an identical pattern as the data. In certain areas, it matches the data perfectly, however it stays more grounded across an equilibrium line compared the to the more fluctating actual data.

We can now look at the diagnostics of our model with the other two variables and the seasonal dummy:

Figure 19.
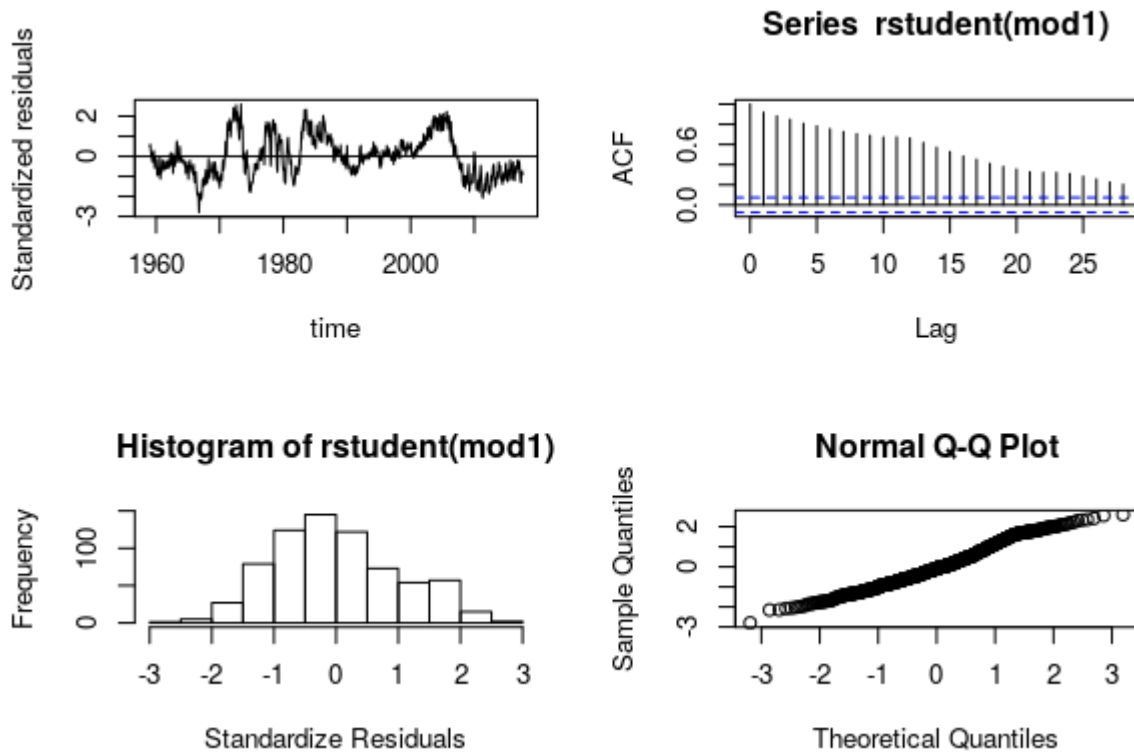
In Figure 19, we notice strong normality in the residuals judging by the shape of the histogram of the residuals and the QQ-plot. The ACF will require further examination with its PACF.

Here I examine the acf and pacf of the residuals for this model:

Figure 20.
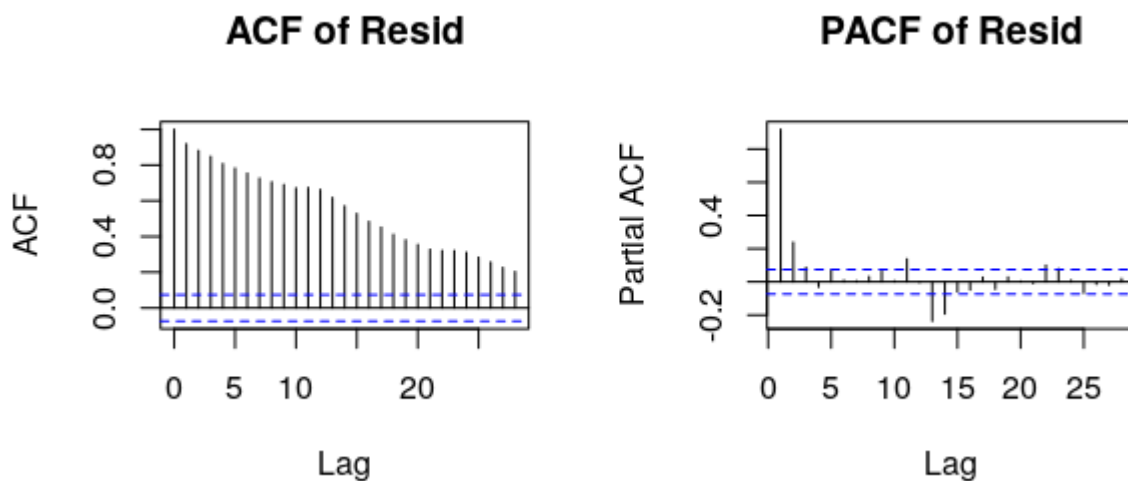
Judging based on Figure 20, there is a signficant lag at the first lag of the PACF indicating an AR(1) model would be a pontential fit to this data. I then fit this model and examine the residuals.

Figure 21.

ACF of Resid
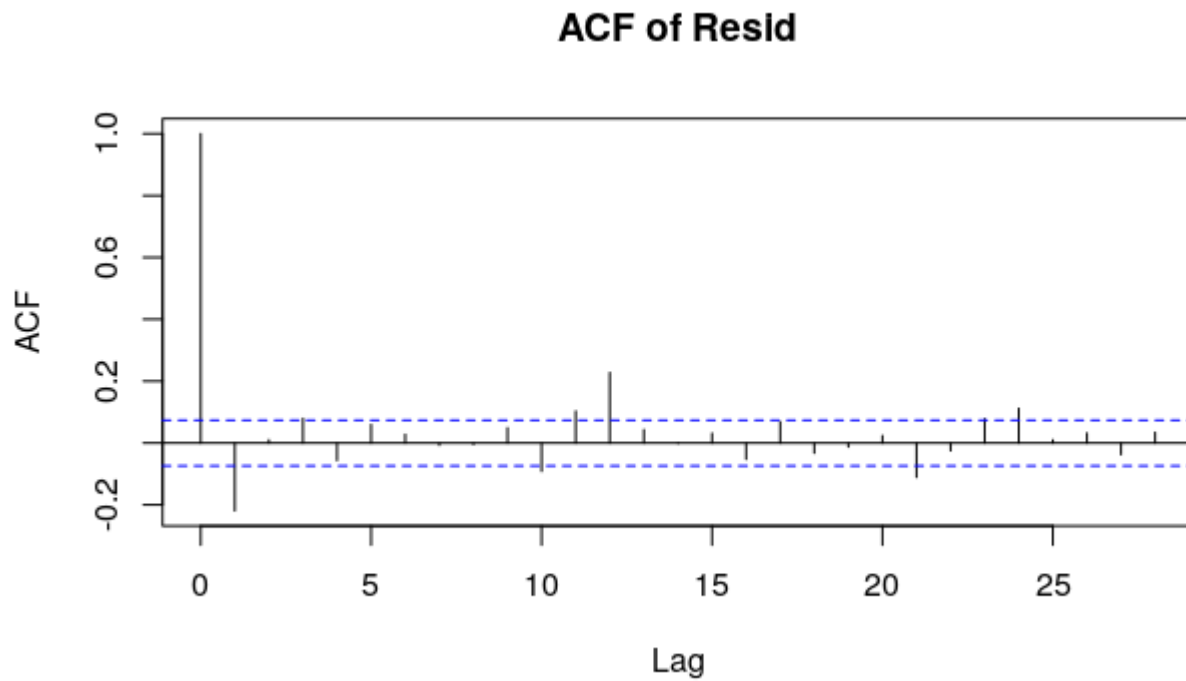
The residuals appear to resemble white noise. To incorporate the informaiton on the correlation of the residuals, I decide to utilize the `gls()` function in R to add a correlation coefficient of 0.9189 which I obtained from looking at the coefficients of the AR(1) model I previously fitted. Let's examine the residuals:

Figure 22.

ACF of Resid of Gls
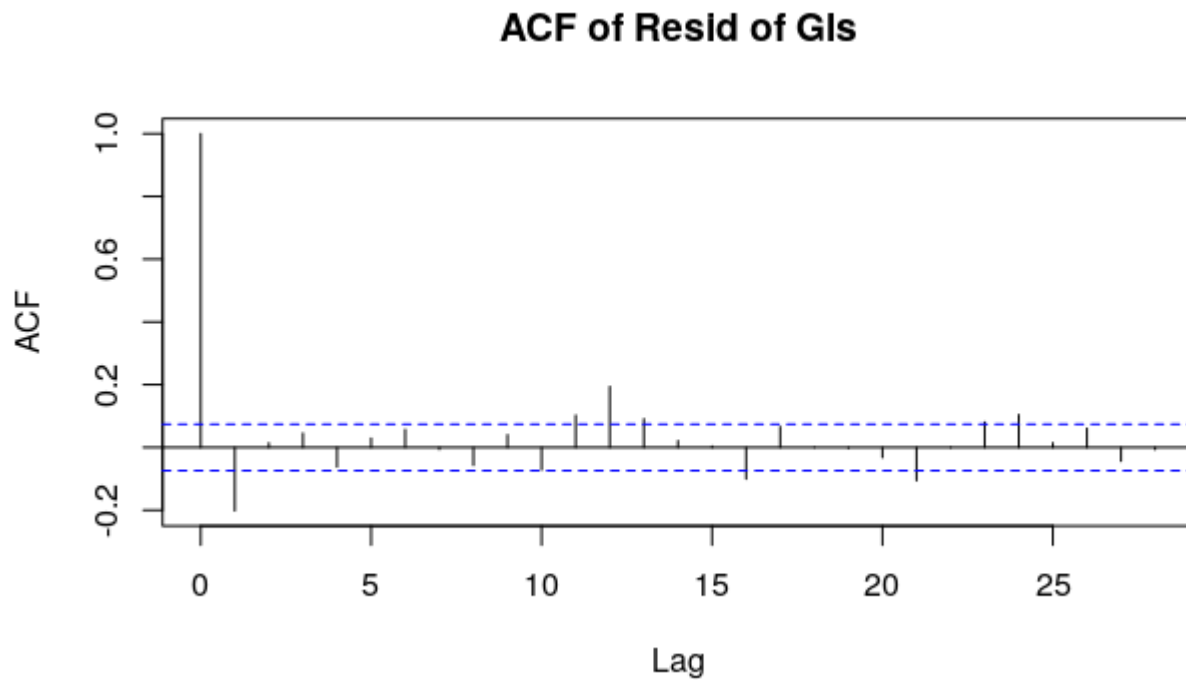
These also appear to be close to white noise.

From here, I decide to forecast and examine the RMSE.

Figure 23.

glsmodel forecast
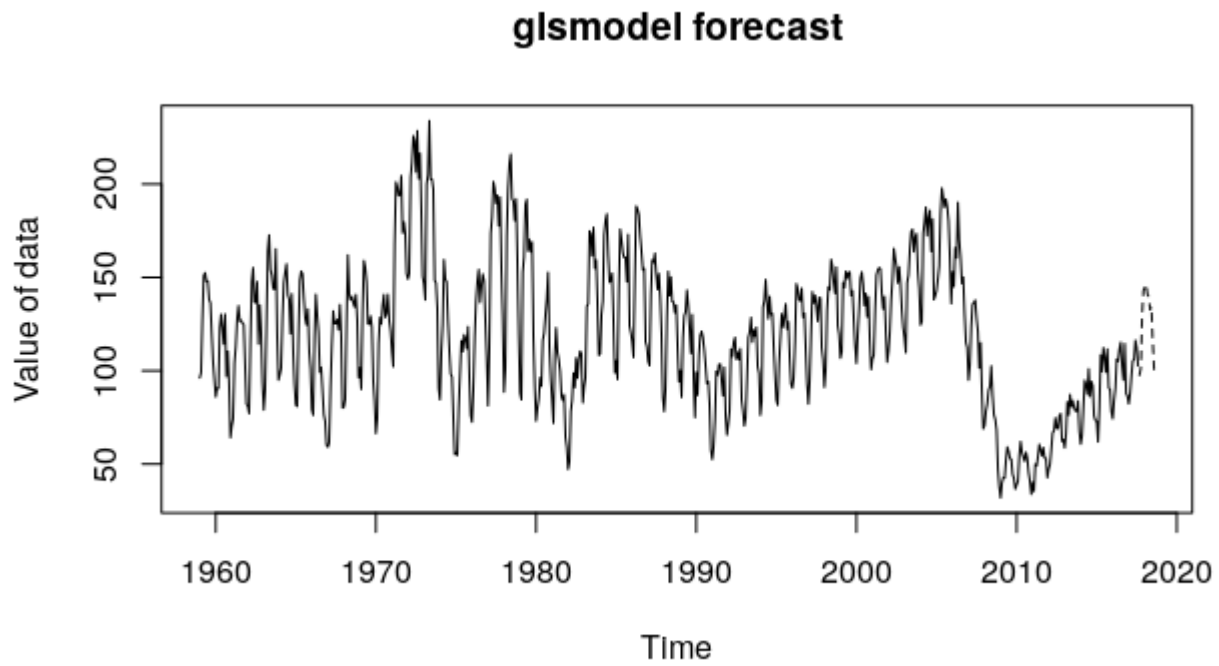
The dashed line in Figure 23 indicates our forecast. It rises in a similar pattern but seems a little high. This particular model achieves an RMSE of 33.32007. To capture any bits of non-stochastic trend, I fit another model with a new trend component added onto the previous model. Let's examine the residuals:

Figure 24.

ACF of Res of New Gls
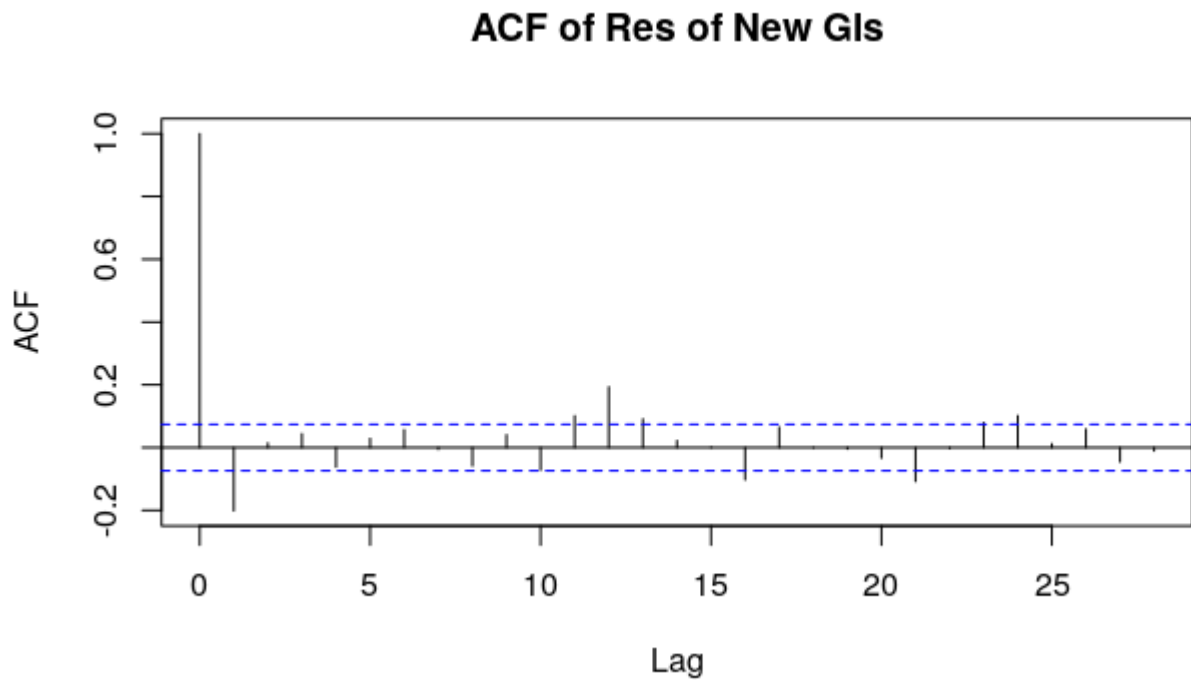
Now, let's forecast once more.

Figure 25.

glsmodel2 forecast
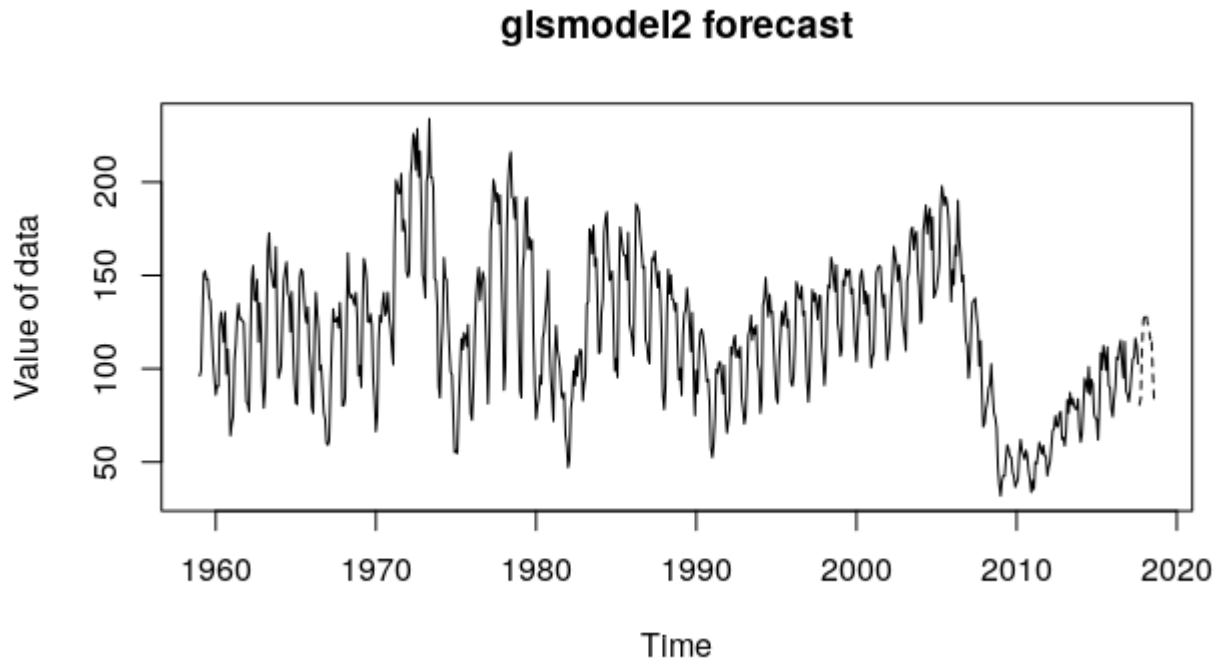
This appears to be further dropped down and more realistic than the previous model which did not account for any such trend. With this new model, I achieve an RMSE of 25.597, a much better result than the previous `gls` model.

6   VAR Model

In this section, I justify the most appropriate VAR model for this data, using all the variables. Next, we observe the acf and difference to help identify a potential model. As depicted in Figure 7, the data is obviously non-stationary and needs to be differenced. In figure 26, we observe the result.

Figure 26.
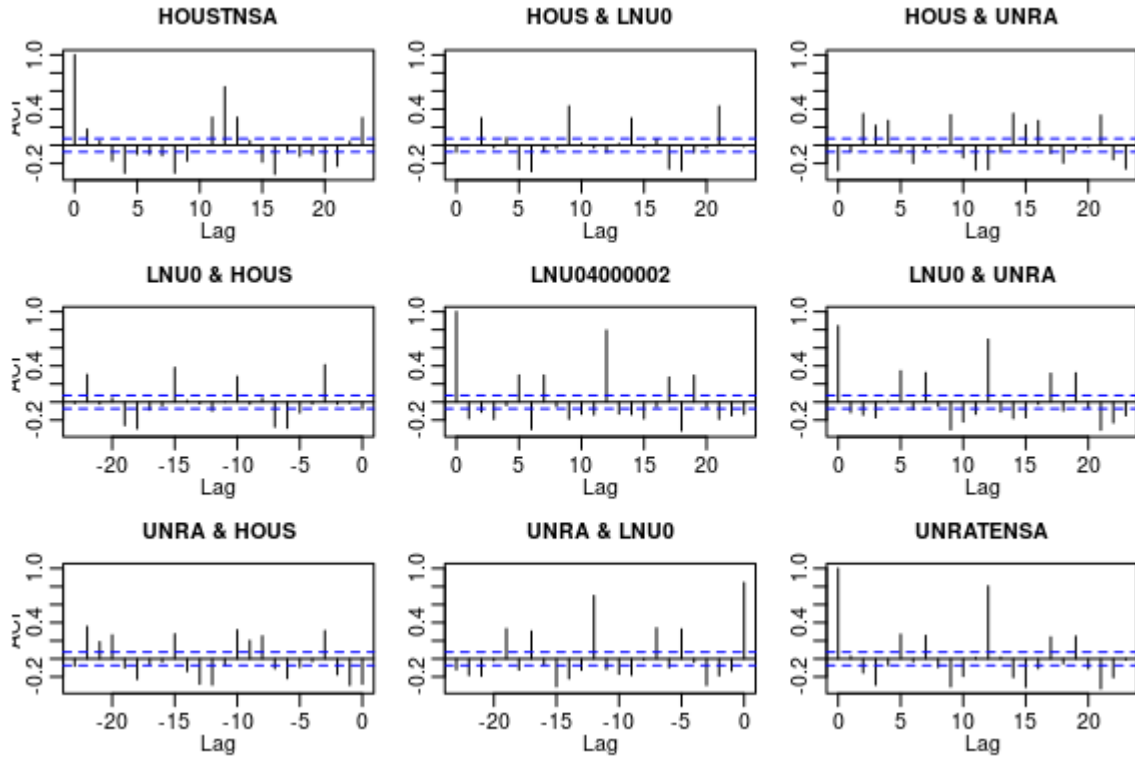
Now the data and relation between the variables are more stationary. In terms of relations of one variable with the other, we notice that there is some pairwise relation between all the variables, all going both ways. We notice that the series at $t-4$ affects the other series at time t then cross correlation drops, suggesting a possible p=4. This is not consistent across all relationships, and in order to verify a correct model fit, we must verify that the residuals are white noise.

When comparing against $p=2, p=3$, and $p=4$, we notice that $p=4$ has white noise residuals as indicated by Figure 27.

Figure 27.
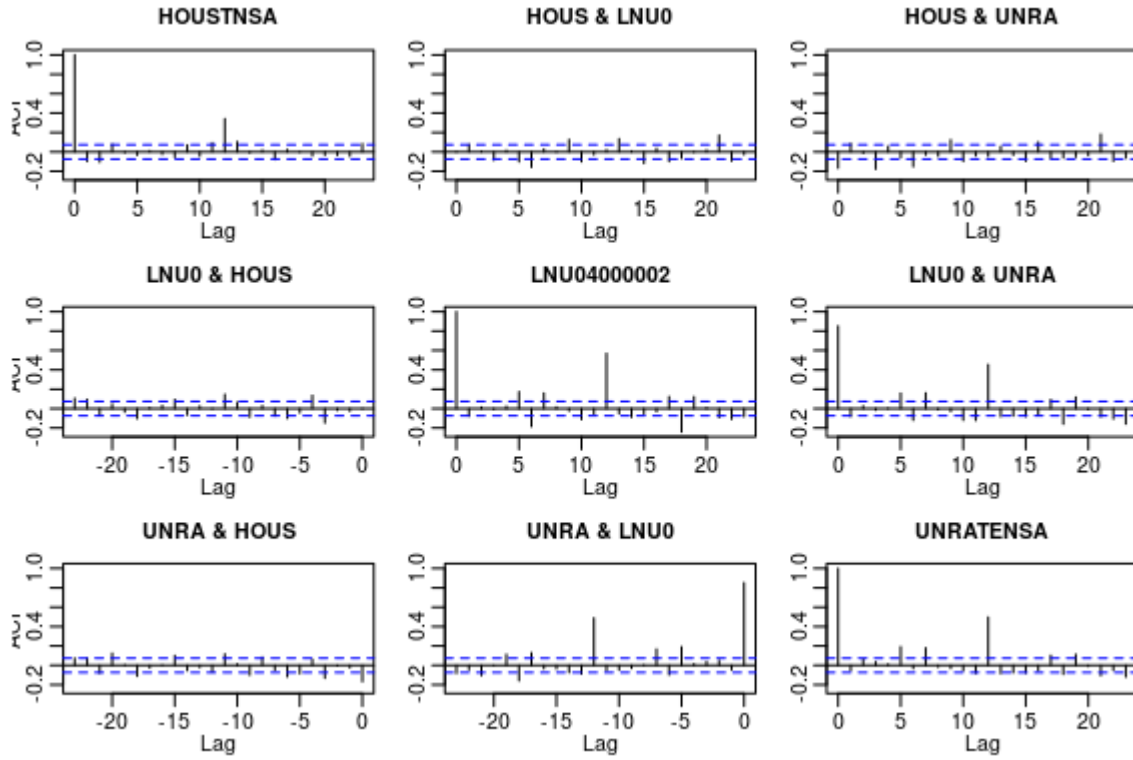
We can now write our final model:

$$hs_t = -0.007^\star t + 0.993 hs_{t-1} + 11.685^\star uw_{t-1} - 13.823 ur_{t-1} + 0.019 hs_{t-2} - 8.533^\star uw_{t-2} + 21.659^\star ur_{t-2} - 0.058 hs_{t-3}$$

$$-26.210^\star uw_{t-3} + 23.398^\star ur_{t-3} + 0.040 hs_{t-4} + 19.203^\star uw_{t-4} - 26.708^\star ur_{t-4}$$

$$uw_t = -0.0002^\star t + 0.001 hs_{t-1} + 0.448^\star uw_{t-1} + 0.417^\star ur_{t-1} - 0.002 hs_{t-2} + 0.087 uw_{t-2} - 0.089 ur_{t-2} + 0.017^\star hs_{t-3}$$

$$+0.021 uw_{t-3} - 0.043 ur_{t-3} + 0.0156^\star hs_{t-4} + 0.283^\star uw_{t-4} - 0.124^\star ur_{t-4}$$

$$ur_t = 3.626 t - 3.860^\star hs_{t-1} - 4.198^\star uw_{t-1} + 1.328^\star ur_{t-1} + 8.835 hs_{t-2} + 2.682 uw_{t-2} - 1.135 ur_{t-2} + 1.409^\star hs_{t-3}$$

$$+1.434 uw_{t-3} - 1.752 ur_{t-3} - 1.150^\star hs_{t-4} + 2.614^\star uw_{t-4} - 4.847 ur_{t-4}$$

Who leads who?
Here, I examine the coefficients of our fitted model to check whether a series lead another. I use the rule: if X has Y as significant, and Y does not have X significant then Y leads X.

<div align="center">

Housing Starts leads: Nothing
Umeployment rate for women leads: Housing starts
Unemployment rate for civilians leads: Housing starts

</div>

Referring back to section 3, the cointegration and unit root tests, we discovered that the variables were cointegrated with each other and each were independent random walks. Since they are cointegrated, this relation is not spurious.

Forecasting
Now, we will forecast the three variables for the next 12 months.

Figure 28.

Forecast of series HOUSTNSA

Forecast of series LNU04000002

Forecast of series UNRATENSA
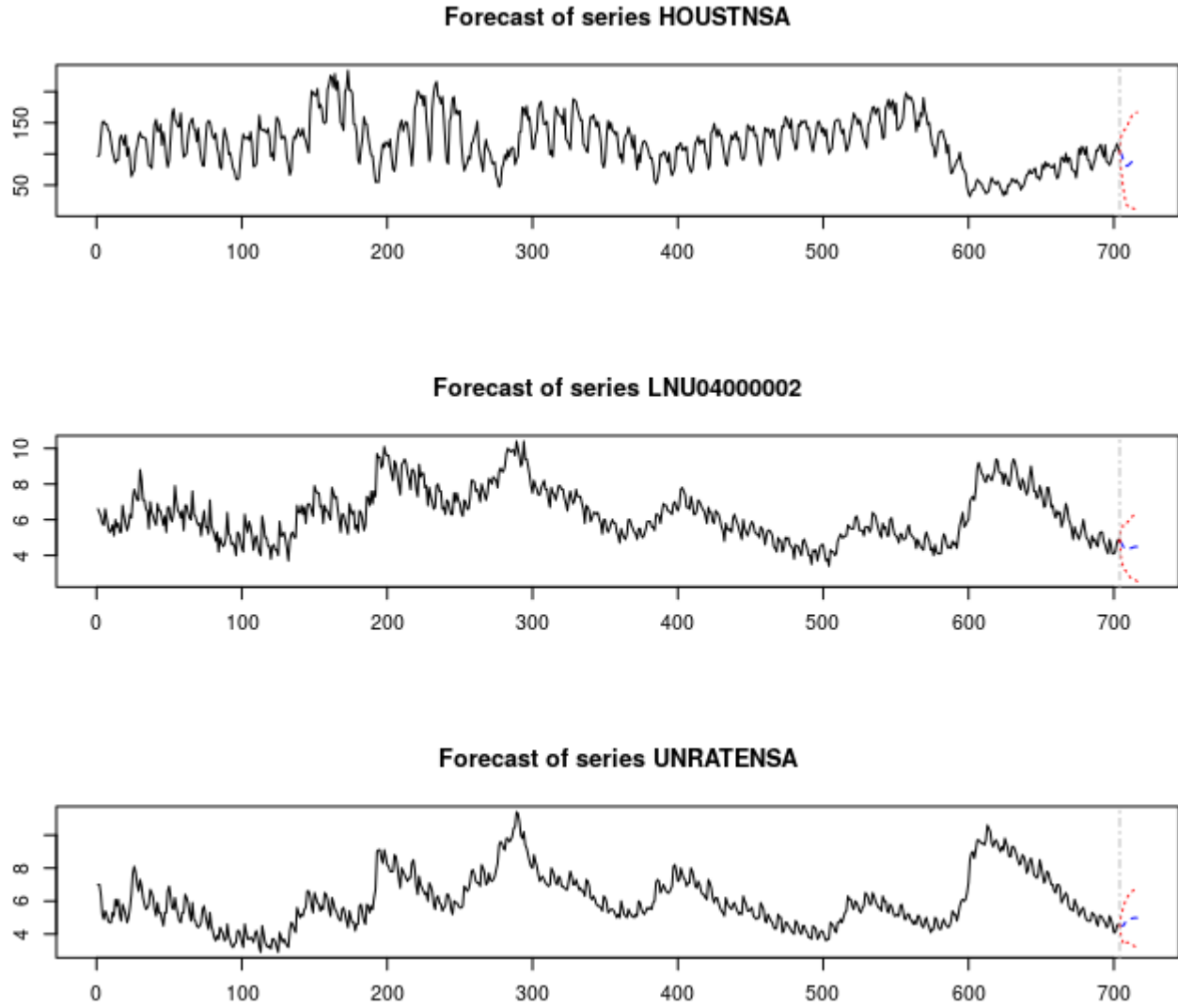
Observing Figure 28, we notice that housing starts is expected to drop and rise again. Unemployment rate for women is expected to drop and stabilize while unemployment for all citizens will gradually increase. By extracting the forecasts just for housing starts, we obtain an RMSE of 20.39057.

Impulse Response Analysis

Here I perform impulse response analysis of a shock to each of the variables separately to study the dynamic relation of the whole estimated system. First, I perform a shock to housing starts and observe the system for the following 5 years.

Figure 29.

Orthogonal Impulse Response from HOUSTNSA
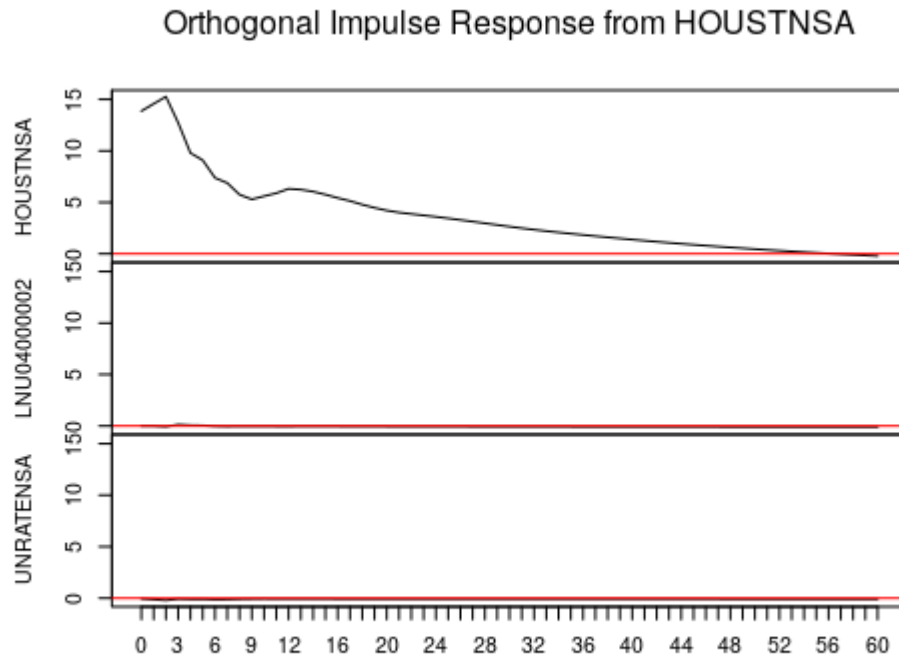
In Figure 29, we notice that a shock to the housing starts leads to negligible effects to the others. Housing starts itself takes an initial rise than gradually falls to equilibrium over the time frame.

Figure 30.

Orthogonal Impulse Response from LNU04000002
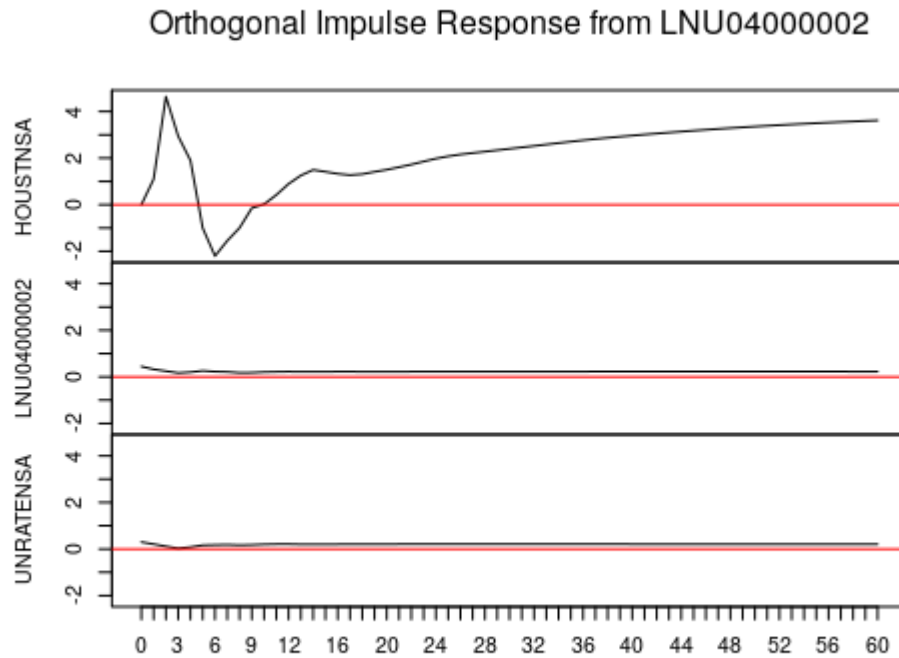
In Figure 30, we notice that a shock to the unemployment rate for women leads to a drastic change in housing starts. Housing starts fluctuates greatly over the equilibrium and then gradually increased over the remaining

22

years resulting in what looks to be a permanent change. The unemployment rate for civilians and women are negligibly affected over time, and share an identical pattern.
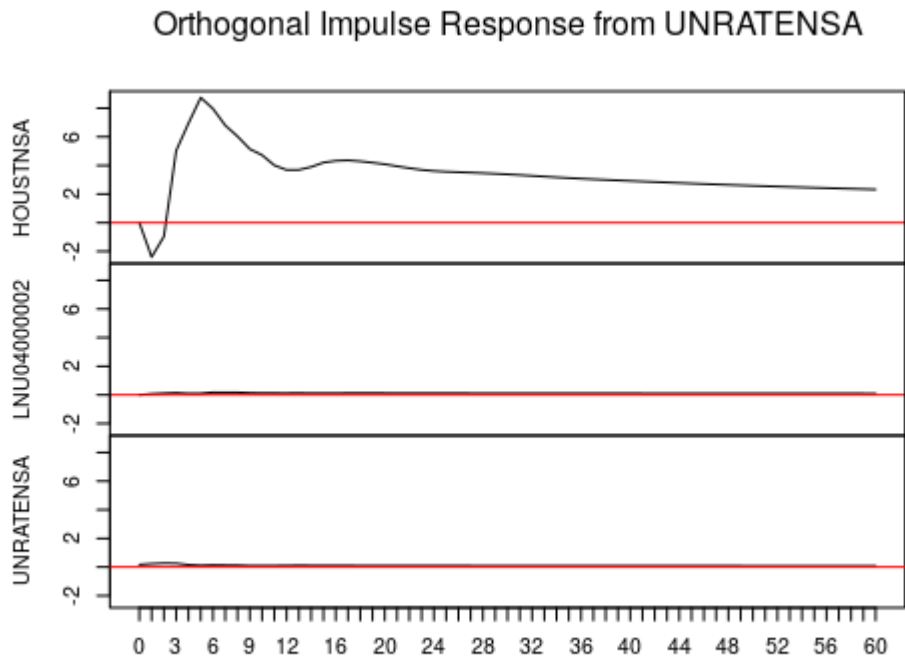
Figure 31.

Orthogonal Impulse Response from UNRATENSA

In Figure 31, we notice that a shock to the unemployment rate for civilians leads to a drastic change in housing starts. Housing initially drops below equilibrium for the first few months then immediately spikes above. After this spike, it begins to gradually fall in the direction of equilibrium. The unemployment rate for civilians and women are negligibly affected over time, and share an identical pattern.

7   Forecasts and Conclusions

Below, I summarize the forecasts from the three models derived in the previous sections for housing starts. Additionally, we present the average of these forecasts and root mean square errors.

Figure 32.

Time Forecasted	SARIMA	Regression	VAR	Avg	Actual Data
Sept 1, 2017	103.81240	80.63850	101.15573	95.20221	104.4
Oct 1, 2017	101.70848	84.23541	93.81887	93.25426	109.6
Nov 1, 2017	88.18002	112.54081	87.43837	96.05307	97.9
Dec 1, 2017	78.49064	126.17320	82.66014	95.77466	81.4
Jan 1, 2018	76.39982	128.02930	80.55905	94.99606	91.6
Feb 1, 2018	79.62674	129.811112	82.19010	97.20932	89.7
Mar 1, 2018	96.05839	124.49648	84.28835	101.61441	107.2
Apr 1, 2018	106.51863	122.78749	86.78125	105.36246	117.5
May 1, 2018	109.31832	115.86656	88.00884	104.39791	123.7
June 1, 2018	112.96804	117.43530	88.46643	106.28992	112.0
July 1, 2018	111.04241	97.75408	88.56794	99.12148	111.9
Aug 1, 2018	104.63629	82.90668	88.11035	91.88444	112.6
RMSE	9.195706	25.59751	20.39057	12.26739	

As we can tell from Figure 32, the SARIMA proved to be the most accurate model in terms of prediction. However, a consensus forecast should be taughtly considered. Consensus forecasts include all possible informationfrom different modeling approaches[2]. If one model tends to overestimate the forecast, it will be balanced out by a forecast which tends to underestimate. A good example of this was that our Regression model tended to overestimate while the VAR model tended to overestimate.

For the most general of cases, I would recommend the consensus forecast due to the points stated above. For this particular case, I would recommend the SARIMA as it proved to be the most accurate a dependable. The RMSE compared to the other models were drastically lower and it was able to correctly understand the structure of housing start's seasonality.

The VAR model tried to take into account the fact that the variables were cointegrated, but it errored on underestimating the forecast. Perhaps, it weighed the actions of the unemployment rates for women and civilians too much which could have decreased the overall housing rate prediction value. The model can be improved. As of now, it is simply a $p = 4$ model, but perhaps further cleaning and model exploration can be done.

The regression model performed the worst compared to the rest of the models, and is not recommended for this particular scenario.

To address the thoughts of economists thinking that housing starts is a leading indicator, I have determined that it does not make sense. As we derived in section 6, housing starts does not lead the unemployment rate for women or civilians. After peeking at the impulse response plot after a shock was applied to housing starts, unemployment rates were not affected whatsoever. On the contrary, when either of the unemployment rates took a shock, housing starts fluctuated greatly to start the short run.

Konner Macias
University of California, Los Angeles

---
[2]Introduction to Time-Series Modeling and Forecasting in Business and Economics. Garynor and Rickey. 1954.